

Improved bounds for the mixing time of the random-to-random insertion shuffle

BEN MORRIS*

CHUAN QIN†

Abstract

We prove an upper bound of $1.5324n \log n$ for the mixing time of the random-to-random insertion shuffle, improving on the best known upper bound of $2n \log n$. Our proof is based on the analysis of a non-Markovian coupling.

Key words: Markov chain, mixing time, non-Markovian coupling.

1 Introduction

How many shuffles does it take to mix up a deck of cards? Mathematicians have long been attracted to card shuffling problems. This is partly because of their natural beauty, and partly because they provide a testing ground for the more general problem of finding the mixing time of a Markov chain, which has applications to computer science, statistical physics and optimization.

Let X_t be a Markov chain on a finite state space V that converges to the uniform distribution. For probability measures μ and ν on V , define the *total variation distance* $\|\mu - \nu\| = \frac{1}{2} \sum_{x \in V} |\mu(x) - \nu(x)|$, and define the ϵ -mixing time

$$T_{\text{mix}}(\epsilon) = \min\{n : \|\Pr(X_t = \cdot) - \mathcal{U}\| \leq \epsilon \text{ for all } x \in V\},$$

where \mathcal{U} denotes the uniform distribution on V .

The random-to-random insertion shuffle has the following transition rule. At each step choose a card uniformly at random, remove it from the deck and then re-insert in to a random position. It has long been conjectured that the mixing time for the random-to-random insertion shuffle on n cards exhibits *cutoff* at a time on the order of $n \log n$. That is, there is a constant c such that for any $\epsilon \in (0, 1)$, the ϵ -mixing time is asymptotic to $cn \log n$. It has further been conjectured (see [2]) that the constant $c = \frac{3}{4}$.

Uyemura-Reyes [8] proved a lower bound of $\frac{1}{2}n \log n$. This was improved by Subag [5] to the conjectured value of $\frac{3}{4}n \log n$. However, a matching upper bound has not been found. Diaconis and Saloff-Coste [3] used comparison techniques to prove a $O(n \log n)$ upper bound. The constant was improved by Uyemura-Reyes [8] and then by Saloff-Coste and Zuniga [6], who proved upper bounds of $4n \log n$ and $2n \log n$, respectively. The main contribution of this paper is to improve the constant in the upper bound to 1.5324. We achieve this via a non-Markovian coupling that reduces the problem of bounding the mixing time to finding the second largest eigenvalue of a certain Markov chain on 9 states. We also use the technique of path coupling (see [1]).

*Department of Mathematics, University of California, Davis. Email: morris@math.ucdavis.edu. Research partially supported by NSF grant CNS-1228828.

†Department of Mathematics, University of California, Davis. Email: cqin@math.ucdavis.edu.

2 Main result

For sequences a_n and b_n , we write $a_n \sim b_n$ if $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 1$ and $a_n \lesssim b_n$ if $\limsup_{n \rightarrow \infty} \frac{a_n}{b_n} \leq 1$. Let P be the transition matrix of the random-to-random insertion shuffle. Define

$$d(t) = \max_x \|P^t(x, \cdot) - \mathcal{U}\| . \quad (1)$$

When the number of cards is n , we write $d_n(t)$ for the value of $d(t)$, and $T_{\text{mix}}^{(n)}(\epsilon)$ for the ϵ -mixing time of the random-to-random insertion shuffle. Our main result is the following upper bound on $T_{\text{mix}}^{(n)}(\epsilon)$.

Theorem 1 *For any $\epsilon \in (0, 1)$ we have $T_{\text{mix}}^{(n)}(\epsilon) \lesssim 1.5324n \log n$.*

We think of a permutation π in S_n as representing the order of a deck of n cards, with $\pi(i) =$ position of card i . Say x and x' are *adjacent*, and write $x \approx x'$, if $x' = (i, j)x$ for a transposition (i, j) . We prove the theorem using a path coupling argument (see [1]) and the following lemma.

Lemma 2 *Suppose $a = 0.6526$. If n is sufficiently large and x and x' are adjacent permutations in S_n , then*

$$\|P^t(x, \cdot) - P^t(x', \cdot)\| \leq e^{-at/n} \quad \text{for all } t > n \log n.$$

The proof of Lemma 2, which uses a non-Markovian coupling, is deferred to Section 3.

Proof of Theorem 1: By convexity of the l^1 -norm, for any state y we have

$$\|P^t(y, \cdot) - \mathcal{U}\| \leq \max_z \|P^t(y, \cdot) - P^t(z, \cdot)\| . \quad (2)$$

Since any permutation in S_n can be written as a product of at most $n - 1$ transpositions, by the triangle inequality the quantity on the righthand side of (2) is at most

$$(n - 1) \max_{x \approx x'} \|P^t(x, \cdot) - P^t(x', \cdot)\| .$$

Let $a = 0.6526$. By Lemma 2, if the number of cards n is sufficiently large we have

$$\max_x \|P^t(x, \cdot) - \mathcal{U}\| \leq (n - 1)e^{-at/n} \quad (3)$$

for all $t > n \log n$. Substituting $1.5324n \log n$ for t in (3), we get

$$d_n(1.5324n \log n) \leq (n - 1)e^{-1.5324a \log n} \quad (4)$$

$$= (n - 1)/n^{1.5324a} \quad (5)$$

$$\rightarrow 0, \quad n \rightarrow \infty, \quad (6)$$

since $1.5324a > 1$.

□

3 Proof of Lemma 2

Recall that we think of a permutation π in S_n as representing the order of a deck of n cards, with $\pi(i)$ = position of card i . Let $M_{i,j} : S_n \rightarrow S_n$ be the operation on permutations that removes the card of label i from the deck and re-inserts it

$$\begin{cases} \text{to the right of the card of label } j & \text{if } i \neq j; \\ \text{to the leftmost position} & \text{if } i = j. \end{cases}$$

We call such operations *shuffles*. If $\langle M_1, \dots, M_k \rangle$ is sequence of shuffles, we write $xM_1M_2 \cdots M_l$ for $M_k \circ M_{k-1} \cdots M_1(x)$.

The transition rule for the random-to-random insertion shuffle can now be stated as follows. If the current state is x , choose a shuffle M uniformly at random (that is, choose a and b uniformly at random and let $M = M_{a,b}$) and move to xM .

We call the numbers in $\{1, \dots, n\}$ *cards*. If shuffle M removes card c from the deck and then re-inserts it, we call M a c -move.

If $\mathcal{P} = \langle M_1, M_2, \dots \rangle$ is a sequence of shuffles, we write $(\mathcal{P}x)_t$ for the permutation $xM_1 \cdots M_t$. Note that if \mathcal{P} is a sequence of independent uniform random shuffles, then $\{(\mathcal{P}x)_t : t \geq 0\}$ is the random-to-random insertion shuffle started at x .

3.1 The coupling

Fix a permutation x and $i, j \in \{1, 2, \dots, n\}$. The aim of this subsection is to define a coupling of the random-to-random insertion shuffle starting from x and $(i, j)x$, respectively.

For positive integers k we will call a sequence $\langle M_1, \dots, M_k \rangle$ of shuffles a k -*path*. For a k -path \mathcal{P} , define the \mathcal{P} -*queue* (or, simply the *queue*) as the following Markov chain $\{Q_t : t = 0, \dots, k\}$ on subsets of cards. Initially, we have $Q_0 = \emptyset$. If the queue at time t is Q_t , and the shuffle at time t is $M_{a,b}$, the next queue Q_{t+1} is

$$\begin{cases} \{i\} & \text{if } a = j; \\ \{j\} & \text{if } a = i; \\ Q \cup \{a\} & \text{if } a \notin \{i, j\} \text{ and } b \in Q_t. \\ Q - \{a\} & \text{otherwise.} \end{cases}$$

We call a shuffle an i -or- j move if it is an i -move or a j -move. For $t \leq k$, we call t a *good time* if

1. t is an i -or- j move;
2. there is a time $t' \in \{t+1, \dots, k\}$ such that
 - (a) t' is the next i -or- j move after t ;
 - (b) the queue is a singleton at time $t' - 1$;
 - (c) the card moved at time t' is different from the card moved at time t .

Let T be the *last* good time in $\{1, \dots, k\}$, with $T = \infty$ if there are no good times, and let $\theta_{i,j}\mathcal{P}$ be the k -path obtained from \mathcal{P} by reversing the roles of i and j in each shuffle before time T (that is, by replacing shuffle $M_{a,b}$ with $M_{\pi(a),\pi(b)}$, where π is a transposition of i and j). Note that $\theta_{i,j}\mathcal{P}$ has i -or- j moves at the same times as \mathcal{P} . Furthermore, since the queue is reset at the times of i -or- j

moves, the $\theta_{i,j}\mathcal{P}$ -queue will have the same values as the \mathcal{P} -queue at all times $t \geq T$. It follows that the last good time of $\theta_{i,j}\mathcal{P}$ is the same as the last good time of \mathcal{P} , and hence $\theta_{i,j}(\theta_{i,j}(\mathcal{P})) = \mathcal{P}$. Since $\theta_{i,j}$ is its own inverse, it is a bijection and hence if \mathcal{P} is a uniform random k -path, then so is $\theta_{i,j}\mathcal{P}$.

Let $x' = (i, j)x$. Let \mathcal{P} be a uniform random k -path, and for t with $0 \leq t \leq k$, define

$$x_t = (\mathcal{P}x)_t \quad x'_t = ((\theta_{i,j}\mathcal{P})x')_t .$$

Let T be the last good time of \mathcal{P} .

Lemma 3 *If $T < k$ then $x_k = x'_k$.*

Proof: Suppose that $T < k$. Note that at any time $t < T$, the permutation $(\mathcal{P}x)_t$ can be obtained from $(\theta_{i,j}\mathcal{P}x')_t$ by interchanging the cards i and j . Let T' be the next i -or- j move after time T . Without loss of generality, there is an i -move at time T and a j -move at time T' . We claim that for times t with $T \leq t < T'$, the permutation x'_t can be obtained from x_t by moving only the cards in Q_t , as shown in the diagram below. (In the diagram, the m th X in the top row represents the same card as the m th X in the bottom row, and Q represents all the cards in Q_t , in any order.)

$$\begin{array}{cccccccccc} x_t : & X & X & X & X & X & X & Q & X & X & X \\ x'_t : & X & X & X & Q & X & X & X & X & X & X \end{array}$$

To see this, note that it holds at time T , when the queue is the singleton $\{j\}$ (since at this time the i 's are placed in the same place), and the transition rule for the queue process ensures that if it holds at time t then it also holds at time $t + 1$. The claim thus follows by induction. This means that at time $T' - 1$ the permutations differ only in the location of card j . That is, they are of the form:

$$\begin{array}{cccccccccc} x_{T'-1} : & X & X & X & X & X & X & j & X & X & X \\ x'_{T'-1} : & X & X & X & j & X & X & X & X & X & X \end{array}$$

Thus at time T' , when card j is removed and then re-inserted into the deck, the two permutations become identical, and they remain identical until time k . \square

Lemma 4 *Suppose $a = 0.6526$. Then for sufficiently large n and $k > n \log n$, we have $\mathbb{P}(T \geq k) \leq e^{-ak/n}$.*

Proof: Consider the Markov chain Y_t defined as follows. The state space is $\{0, 1, \dots\} \cup \infty$. The chain starts in state ∞ and remains there until the first i -or- j move. From this point on, the value of Y_t is the size of the queue, until the first time that either

1. card i is moved when the queue is $\{i\}$, or
2. card j is moved when the queue is $\{j\}$.

At this point Y_t moves to state 0, which is an absorbing state. Note that $T < k$ exactly when $Y_k = 0$.

For $l = 1, 2, \dots$, define

$$q(l) = \begin{cases} \frac{1}{n} & \text{if } l = 1, \\ \frac{3n-1}{n^2} & \text{if } l = 2, \\ \frac{(l-1)(n-l+1)}{n^2} & \text{if } l \geq 3; \end{cases}$$

and define

$$p(l) = \begin{cases} \frac{n-2}{n^2} & \text{if } l = 1, \\ \frac{2n-6}{n^2} & \text{if } l = 2, \\ \frac{l(n-l-1)}{n^2} & \text{if } l \geq 3. \end{cases}$$

The transition rule for Y_t can be described as follows. If the current state is 0, the next state is 0. If the current state is ∞ the next state is

$$\begin{cases} 1 & \text{with probability } \frac{2}{n}; \\ \infty & \text{with probability } \frac{n-2}{n}. \end{cases}$$

If the current state is $l \in \{1, 2, \dots\}$, the next state is

$$\begin{cases} l-1 & \text{with probability } q(l); \\ l+1 & \text{with probability } p(l); \\ 1 & \text{with probability } \frac{2}{n}, \text{ if } l \geq 3; \\ l & \text{with the remaining probability.} \end{cases}$$

Let \tilde{Y}_t be the Markov chain on $\{0, 1, \dots, 7\} \cup \infty$ obtained from Y_t by replacing transitions to state 8 with transitions to ∞ . That is, if K and \tilde{K} denote the transition matrices of Y_t and \tilde{Y}_t , respectively, then

$$\tilde{K}(l, m) = \begin{cases} K(l, m) & \text{if } m \in \{0, 1, \dots, 7\}; \\ K(7, 8) & \text{if } l = 7 \text{ and } m = \infty. \end{cases}$$

The possible transitions of Y_t and \tilde{Y}_t are indicated by the graph in Figure 1. We claim that if we start with $\tilde{Y}_0 = Y_0 = \infty$ then the distribution of \tilde{Y}_t stochastically dominates the distribution of Y_t for all t . To see this, note that Y_t changes state with probability less than $\frac{1}{2}$ at each step, and when it changes state, it either makes a ± 1 move or it transitions to 1. Since for $m \in \{1, 2, \dots\} \cup \infty$, the transition probability $K(m, 1)$ is decreasing in m , it follows that Y_t is a monotone chain. The claim follows since \tilde{Y}_t is obtained from Y_t by replacing moves to 8 with moves to the (larger) state of ∞ .

Let \tilde{K}_n be the value of the matrix \tilde{K} when the number of cards is n . Define $B_n = \tilde{K}_n - I$, where

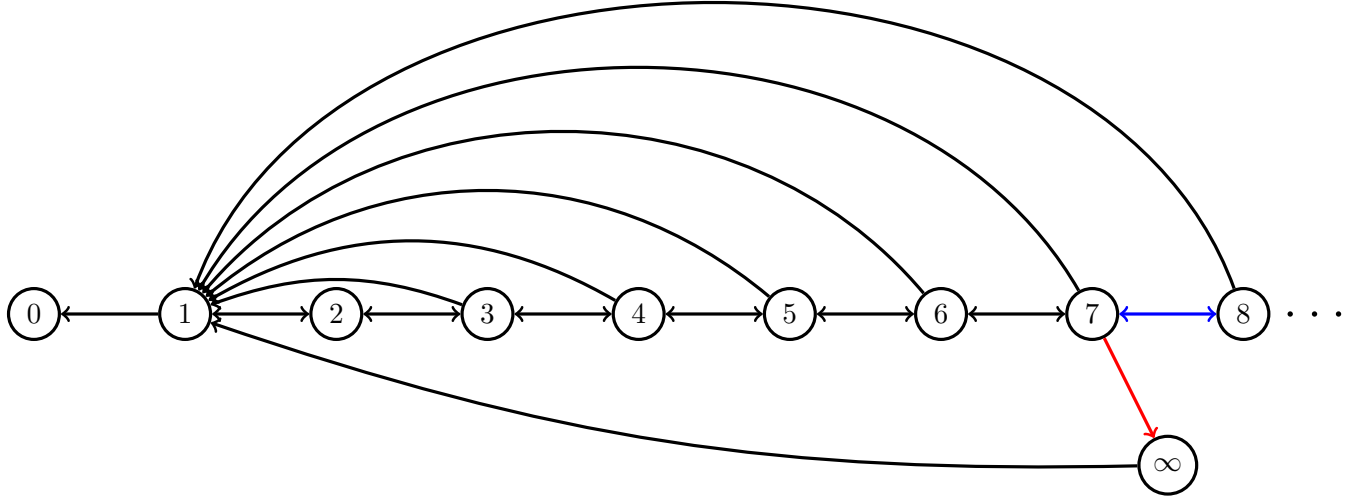


Figure 1: Graph indicating the possible transitions of Y_t and \tilde{Y}_t . (The blue edge indicates a possible transition of Y_t and the red edge indicates a possible transition of \tilde{Y}_t .)

I is the identity matrix. A straightforward calculation shows that $nB_n \rightarrow C$ as $n \rightarrow \infty$, where

$$C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & -5 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 2 & -7 & 3 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 3 & -9 & 4 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 4 & -11 & 5 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 5 & -13 & 6 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 6 & -15 & 7 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & -2 \end{bmatrix}.$$

For matrices A we write $\lambda(A)$ for the *second largest* eigenvalue of A . By the relationship between l^2 -norm and eigenvalues, we have

$$\left(\mathbb{P}(\tilde{Y}_k = \infty)^2 + \sum_{l=1}^7 \mathbb{P}(\tilde{Y}_k = l)^2 \right)^{1/2} \leq \left(\lambda(\tilde{K}_n) \right)^k \quad (7)$$

$$= [1 + \lambda(B_n)]^k \quad (8)$$

$$\leq \exp(k\lambda(B_n)). \quad (9)$$

Since Y_t is stochastically dominated by \tilde{Y}_t ,

$$\mathbb{P}(Y_k > 0) \leq \mathbb{P}(\tilde{Y}_k = \infty) + \sum_{l=1}^7 \mathbb{P}(\tilde{Y}_k = l) \leq \sqrt{8} \left(\mathbb{P}(\tilde{Y}_k = \infty)^2 + \sum_{l=1}^7 \mathbb{P}(\tilde{Y}_k = l)^2 \right)^{1/2}, \quad (10)$$

where the second inequality is by Cauchy-Schwarz. Let J_n denote the matrix obtained by deleting the first column and the first row from $I + (1/16)(nB_n)$, and note that $1 + \lambda(nB_n)/16$ is the largest eigenvalue of J_n . Since J_n is a sub-stochastic matrix for all n , it is well known that the largest eigenvalue of J_n converges to the largest eigenvalue of the entry-wise limit of J_n as $n \rightarrow \infty$. This implies that $\lim_{n \rightarrow \infty} \lambda(nB_n) = \lambda(C)$. Numerical calculations show that $\lambda(C) < a := -0.6526$. Therefore, there exists some constant $\delta > 0$ such that

$$k\lambda(B_n) = \frac{k}{n}\lambda(nB_n) \leq -(a + \delta)k/n$$

for sufficiently large n . This combined with (9) and (10) proves that

$$\mathbb{P}(T \geq k) = \mathbb{P}(Y_k > 0) \leq \sqrt{8} \exp(-(a + \delta)k/n) \leq \exp(-ak/n)$$

for sufficiently large n and $k > n \log n$. □

Proof of Lemma 2: Recall that for any two probability measures μ and ν on a probability space Ω , we have

$$\|\mu - \nu\| = \min\{\mathbb{P}(X \neq Y) : (X, Y) \text{ is a coupling of } \mu \text{ and } \nu\}.$$

The main lemma then follows immediately from Lemma 3 and Lemma 4. □

References

- [1] R. Bubley and M. Dyer, Path Coupling: A technique for proving rapid mixing in Markov Chains, Proceedings of the 38th Annual Symposium on Foundation of Computer Science, pp. 223-231, 1997.
- [2] P. Diaconis, Mathematical developments from the analysis of riffle shuffling, In Groups, Combinatorics, Geometry (Durham 2001), 73-97, 2001.
- [3] P. Diaconis and L. Saloff-Coste, Comparison techniques for random walks on finite groups, Ann. Probab. 21, 2131-2156, 1993.
- [4] D. Levin, Y. Peres and E. Wilmer, Markov Chains and mixing time, American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.
- [5] E. Subag, A Lower Bound for the Mixing Time of the Random-to-Random Insertions Shuffle, Electron. J. Probab. 18, 1-20, 2012.
- [6] L. Saloff-Coste and J. Zuniga, Refined estimates for some basic random walks on the symmetric and alternating groups, Latin American Journal of Probability and Mathematical Statistics 4, 359-392, 2008.
- [7] L. Saloff-Coste and J. Zuniga, Convergence of some time inhomogeneous Markov chains via spectral techniques. Stochastic Processes and their Applications 117, 961-979, 2007.
- [8] J. Uyemura-Reyes, Random Walk, semi-direct products, and card shuffling, Ph.D. Thesis, Stanford University, 2002.